Ann. rheum. Dis. (1957), **16**, 485.

RADIOLOGICAL ASSESSMENT OF RHEUMATOID ARTHRITIS*

BY

J. H. KELLGREN AND J. S. LAWRENCE

From the Rheumatism Research Centre, Manchester University, and the Empire Rheumatism Council Field Unit

Introduction

In a previous paper (Kellgren and Bier, 1956), a study was made of inter- and intra-observer differences in grading hand films for rheumatoid arthritis during the course of therapeutic trials and field surveys. This study indicated that standards used so far in grading *x*-ray changes for rheumatoid arthritis in field surveys in Leigh and South Wales did not differ greatly from the mean values given by physicians and radiologists in Britain.

Before attempting to produce standard films for use in future surveys, it was decided to seek the help of a number of workers in different parts of the world, so that any national bias in assessing such changes could be avoided. Advantage was, therefore, taken of the presence in Toronto of experienced rheumatologists from many countries on the occasion of the XI International Congress of Rheumatic Diseases in June, 1957. During this Congress a series of nineteen hand *x*-rays with varying degrees of rheumatoid arthritis and osteo-arthrosis were exhibited in the form of a radiological quiz, and members of the Congress were asked to grade these for osteo-porosis and rheumatoid erosions and to give an overall grading for rheumatoid arthritis and osteo-arthrosis (degenerative joint disease). Each of these features was to be graded into five categories:

0 = None
1 = Doubtful
2 = Minimal
3 = Moderate
4 = Severe

Erosions were to include narrowing of joint space due to erosion of cartilage as well as localized erosions of bone. Simple bone cysts and narrowing of cartilage due to degenerative joint disease were to be exempted. It was, however, recognized that there would be difficulties in distinguishing between narrowing due to rheumatoid arthritis and degenerative joint disease. The films used for this quiz were taken largely from the Leigh population survey of persons between the ages of 55 and 64 (Kellgren and Lawrence, 1956), but six films from one of the therapeutic trials conducted by the Medical Research Council and Nuffield Foundation Joint Committee (1954) were also included. The films were chosen to illustrate all grades of severity of the four factors. The Leigh films had been read by two observers in consultation, together and separately, the trial films by six observers independently; one of whom also took part in the Leigh readings. To facilitate the grading of osteo-porosis a standard bone had been used when taking the *x* rays.

Altogether 62 persons took part in the radiological quiz, but the answers sent in by two of them were incomplete so that this analysis is based on the observations of sixty persons. The country of origin of those taking part is as follows:

United States of America	40
United Kingdom	7
Continent of Europe	6
Canada	4
Australia	2
Brazil	1

Results

Comparison of Original British Readings with Mean Averages for All Readings

Osteo-Porosis.—There was substantial agreement. In ten of the nineteen films the average grading was identical, in three the original readings were graded higher and in six lower than those of the Congress observers. In no film was there a difference of two grades (Table I).

COMPARISON OF AVERAGE TORONTO WITH ORIGINAL BRITISH READINGS

TABLE I

OSTEO-POROSIS

Original British Readings	0	1	2	3	4	Total
4	—	—	—	—	2	2
3	—	—	2	—	—	2
2	—	1	3	—	—	4
1	—	2	—	—	—	2
0	3	6	—	—	—	9
Total	3	9	5	0	2	19

Average Readings at Toronto

* Results of a radiological quiz held at the IX International Congress of Rheumatic Diseases at Toronto in June, 1957.

Erosions.—There was agreement in seven of the nineteen films. Eight received a higher grading by the original British observers and four a lower. In two films there was a two-grade difference, the original readings being higher in one and lower in one (Table II). A higher grading for erosions was given by Congress observers chiefly when osteo-arthrosis was present and probably represents a different interpretation of the term erosion, so that degenerative narrowing of the cartilage, and possibly also bone cysts, came to be included. As, however, these particular films were not in most instances given a definite grading for rheumatoid arthritis, this does not invalidate the conclusions on rheumatoid arthritic grading which follow.

TABLE II

EROSIONS

Original British Readings	Total	2	6	6	3	2	19
	4	—	—	—	2	2	4
	3	—	1	3	1	—	5
	2	—	2	—	—	—	2
	1	—	2	2	—	—	4
	0	2	1	1	—	—	4
	Score	0	1	2	3	4	Total

Average Readings at Toronto

Rheumatoid Arthritis.—The same grading was given on nine of the films, a higher grading by the original British observers on nine and a lower grading on only one. In none was there a difference of more than one grade (Table III). The correlation coefficient ($r = +0{\cdot}91$) is seven times the standard error and indicates a very satisfactory correlation despite the tendency for the original observers to grade higher.

TABLE III

RHEUMATOID ARTHRITIS

Original British Readings	Total	5	6	3	3	2	19
	4	—	—	—	2	2	4
	3	—	—	2	1	—	3
	2	—	3	1	—	—	4
	1	2	2	—	—	—	4
	0	3	1	—	—	—	4
	Score	0	1	2	3	4	Total

Average Readings at Toronto
$r = \pm 0{\cdot}91$

Osteo-Arthrosis.—The original British observers agreed closely with the Congress observers. In ten the average grading was identical, in six a higher average grading was given by the original observers and in three a lower grading. There were no two-grade differences (Table IV).

TABLE IV

OSTEO-ARTHROSIS

Original British Readings	Total	7	9	0	3	0	19
	4	—	—	—	1	—	1
	3	—	—	—	2	—	2
	2	—	3	—	—	—	3
	1	2	3	—	—	—	5
	0	5	3	—	—	—	8
	Score	0	1	2	3	4	Total

Average Readings at Toronto

The average readings for all films are shown in Table V. These figures confirm that, except for osteo-porosis, the original observers tended to grade higher, but only in their readings of rheumatoid arthritis was the difference appreciable, even then averaging less than half a grade.

TABLE V

AVERAGE READINGS AT TORONTO CONGRESS COMPARED WITH ORIGINAL BRITISH READINGS

Readings	Osteo-Porosis	Erosions	Rheumatoid Arthritis	Osteo-Arthrosis
Original British	1·3	2·0	1·9	1·1
Average Toronto Congress	1·4	1·8	1·6	0·9

Regional Comparisons

Average readings (for all films) by workers from different countries are shown in Table VI where they are compared with the original British readings.

TABLE VI

AVERAGE GRADINGS FOR ALL FILMS BY WORKERS FROM DIFFERENT COUNTRIES

Country	Number of Readers	Osteo-Porosis	Erosions	Rheumatoid Arthritis	Osteo-Arthrosis
United States of America and Brazil ..	41	1·5	1·8	1·6	1·1
United Kingdom ..	7	1·0	1·8	1·7	1·1
Canada	4	1·6	1·8	1·9	1·2
Continent of Europe ..	6	1·4	2·2	1·7	0·8
Australia	2	1·8	2·1	1·5	1·3
Original British Readings	6	1·3	2·0	1·9	1·1

The tendency is for rheumatoid arthritis to be graded lower by the American and Australian observers, and higher by Canadians and the original British group, the Continental European observers occupying an intermediate position. In the grading of osteo-arthrosis and of erosions there is little regional difference. Osteo-porosis was read highest by the Australians and lowest by the British.

TABLE VII

RHEUMATOID ARTHRITIS

COMPARISON OF TWO GROUPS OF TORONTO OBSERVERS

United States of America	Total	4	6	3	4	2	19
	4	—	—	—	—	2	2
	3	—	—	—	3	—	3
	2	—	—	2	1	—	3
	1	1	5	1	—	—	7
	0	3	1	—	—	—	4
	Score	0	1	2	3	4	Total

Other Observers at Toronto
$r = +0{\cdot}94$

A comparison of the gradings of rheumatoid arthritis on individual films between observers from the United States of America and all the others at Toronto (Table VII) shows a highly satisfactory degree of correlation ($r = +0{\cdot}94$).

Data on Individual Films

To gain further information on grading, four films were chosen for detailed study (Figs 1-4). The rheumatoid grading, according to general consensus of opinion, was: Severe (Film N), Moderate (Film O), Minimal (Film Q), Nil (Film G).

Rheumatoid Arthritis.—Readings by individual observers on these first four films are shown in Fig. 5 (overleaf).

On Film N, which shows the most severe changes, there was almost universal agreement, 87 per cent. of observers placing it in Grade 4 and the remainder in Grade 3. Similarly, there was fairly uniform agreement on Film G which was considered free of rheumatoid arthritis by the original observers and also by 78 per cent. of the Toronto observers,

Fig. 1.—Film N, showing severe rheumatoid arthritis. The majority gradings were Grade 4 porosis, Grade 4 erosions, and Grade 4 rheumatoid arthritis.

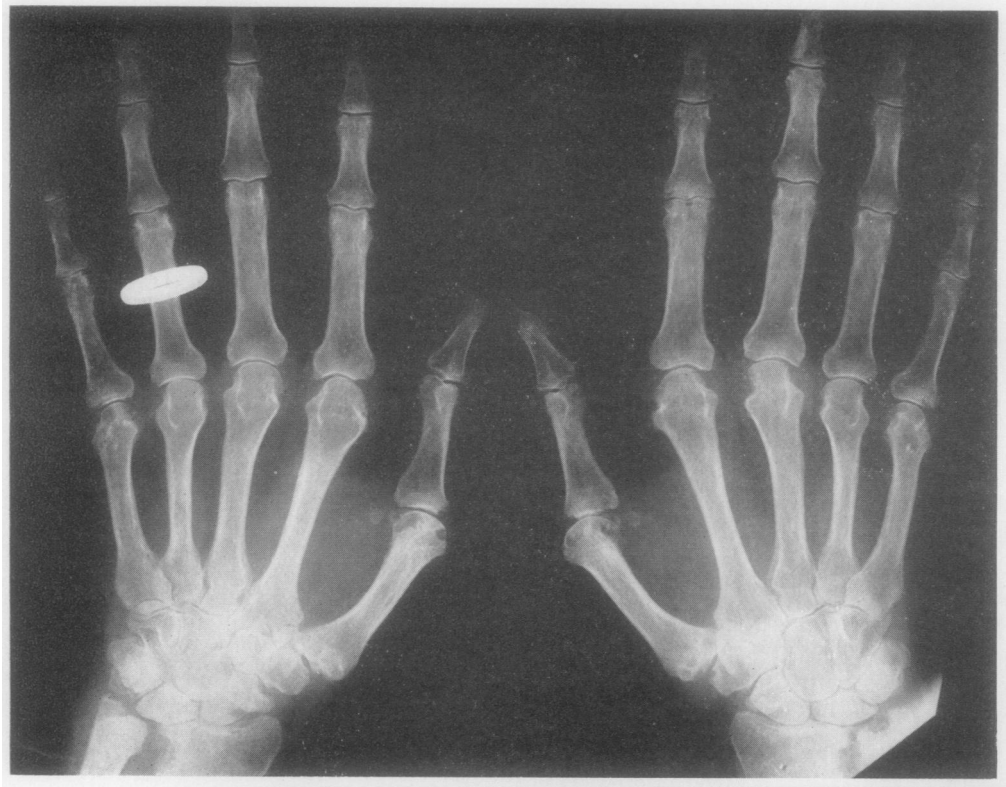


Fig. 2.—Film O, showing moderate rheumatoid arthritis. The majority gradings were Grade 2 porosis, Grade 3 erosions, and Grade 3 rheumatoid arthritis.

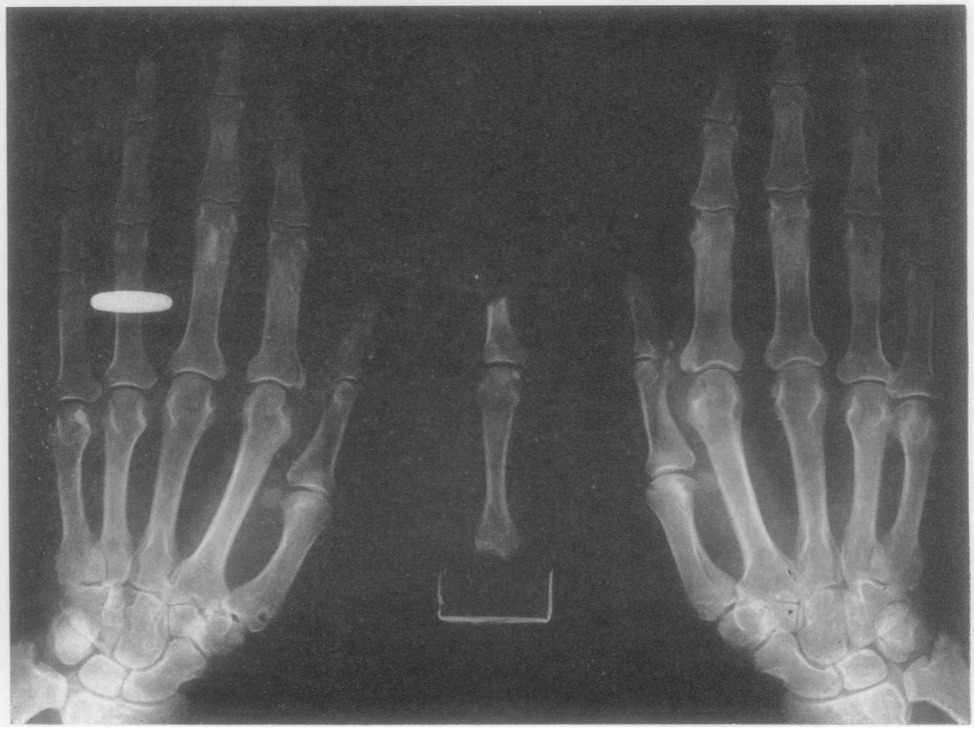


Fig. 3.—Film Q showing minimal rheumatoid arthritis. The majority gradings were Grade 2 porosis, Grade 2 erosions, and Grade 2 rheumatoid arthritis.

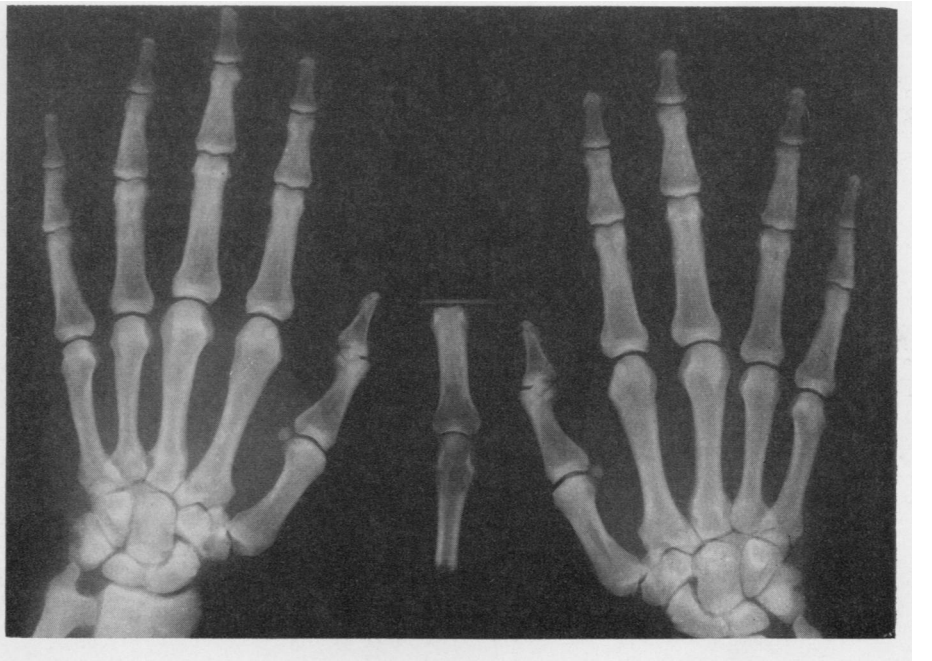
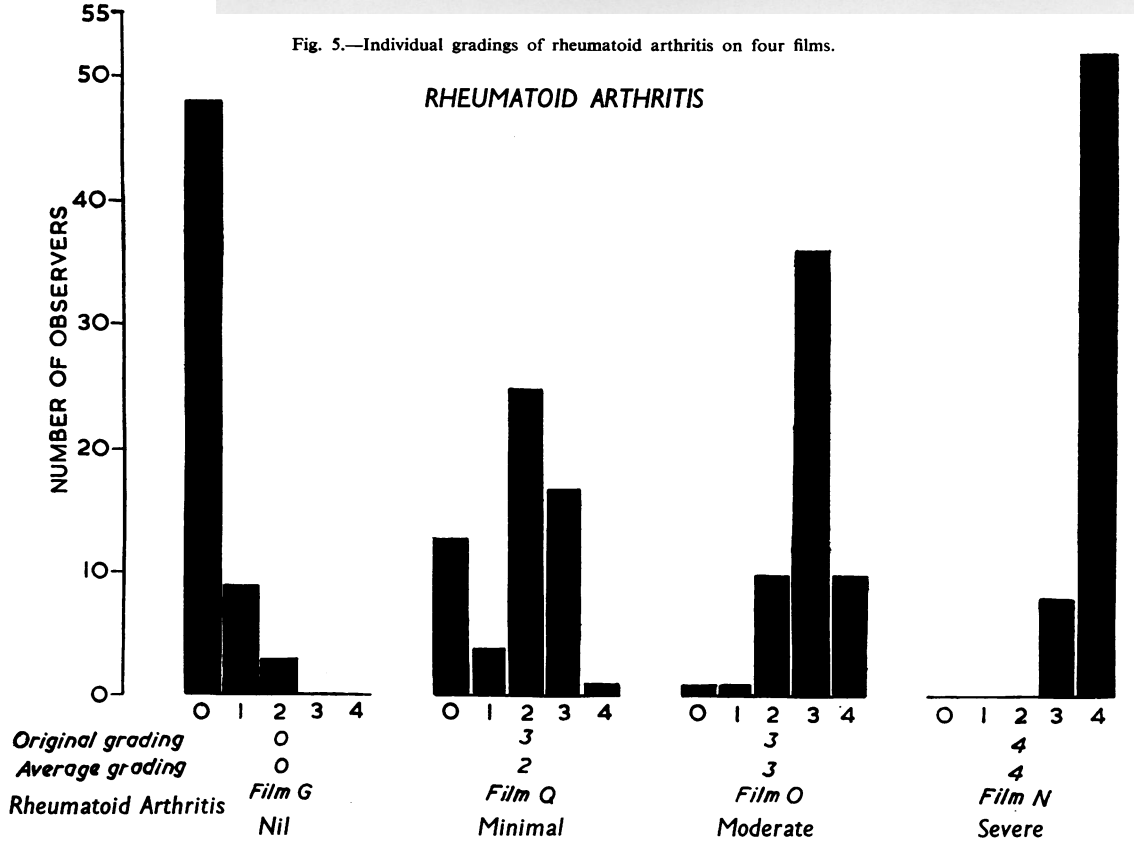


Fig. 4.—Film G, showing a normal subject with no rheumatoid arthritis. The majority gradings were Grade 0 porosis, Grade 0 erosions, and Grade 0 rheumatoid arthritis.

Fig. 5.—Individual gradings of rheumatoid arthritis on four films.



though three persons (5 per cent.) gave it Grade 2 and nine (15 per cent.) Grade 1. On Film O, originally considered to show moderate changes, agreement was less marked, 60 per cent. giving Grade 3, but one observer giving Grade 0 and another Grade 1, these two preferring a diagnosis of osteo-arthrosis.

With Film Q, which had only minimal evidence of rheumatoid arthritis, there was considerable scatter, 42 per cent. Grade 2, 28 per cent. Grade 3, 22 per cent. nil, and 2 per cent. Grade 4.

Osteo-Porosis (Fig. 6).—There was a considerable consensus of opinion on Films G and N with no porosis and severe porosis respectively, but in Films Q and O, in which the osteo-porosis is less severe, there were wide differences in grading. Particularly is this so in Film P (Fig. 7), in which roughly the same number of observers gave gradings 0, 1, 2, or 3. Difficulty appears to have arisen in this film because the osteo-porosis was limited to the neighbourhood of a few joints and general porosis was lacking.

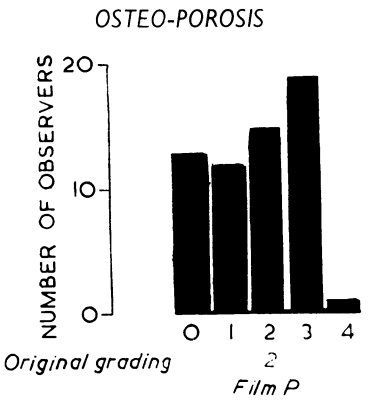


Fig. 7.—Individual gradings of osteo-porosis on Film P, in which the porosis was limited to the neighbourhood of a few joints.

Erosions (Fig. 8, opposite).—Films G and N were again uniformly graded, 93 per cent. giving Grade 0 in G and 79 per cent. Grade 4 in N. In Films Q and O, from patients with intermediate degrees of

Fig. 6.—Individual gradings of osteo-porosis on four films.

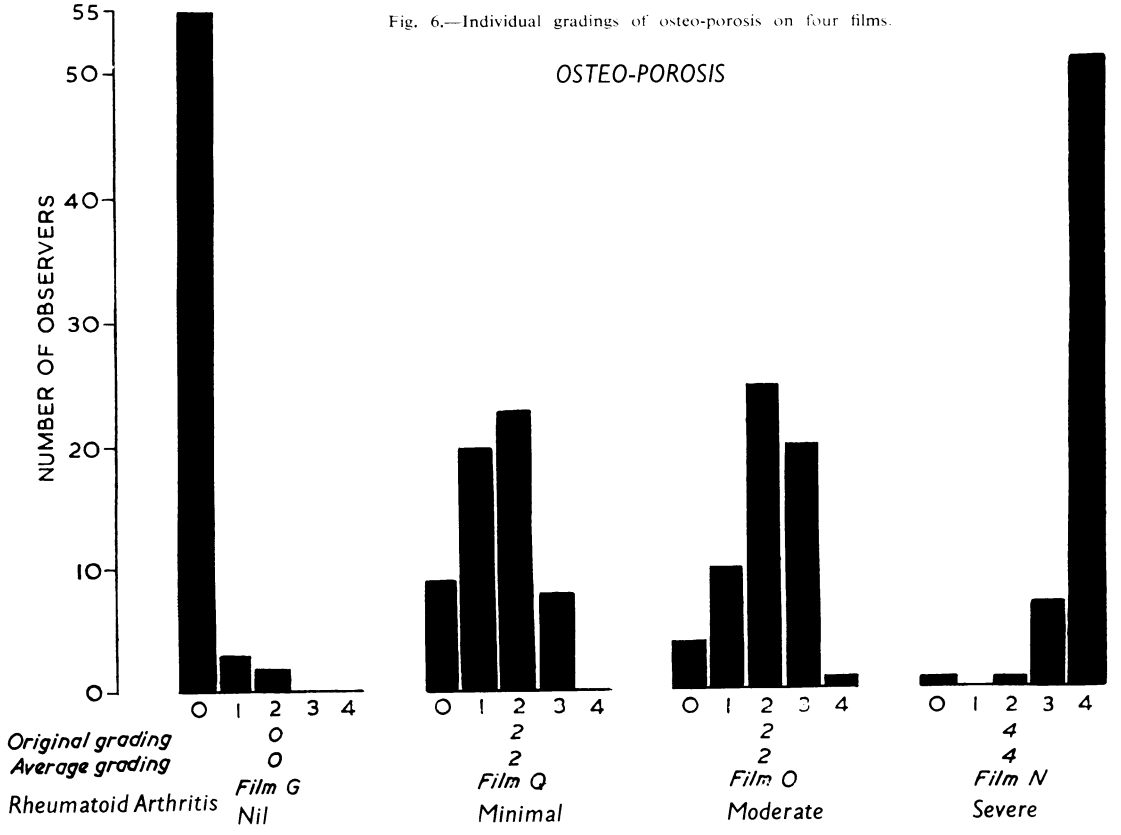
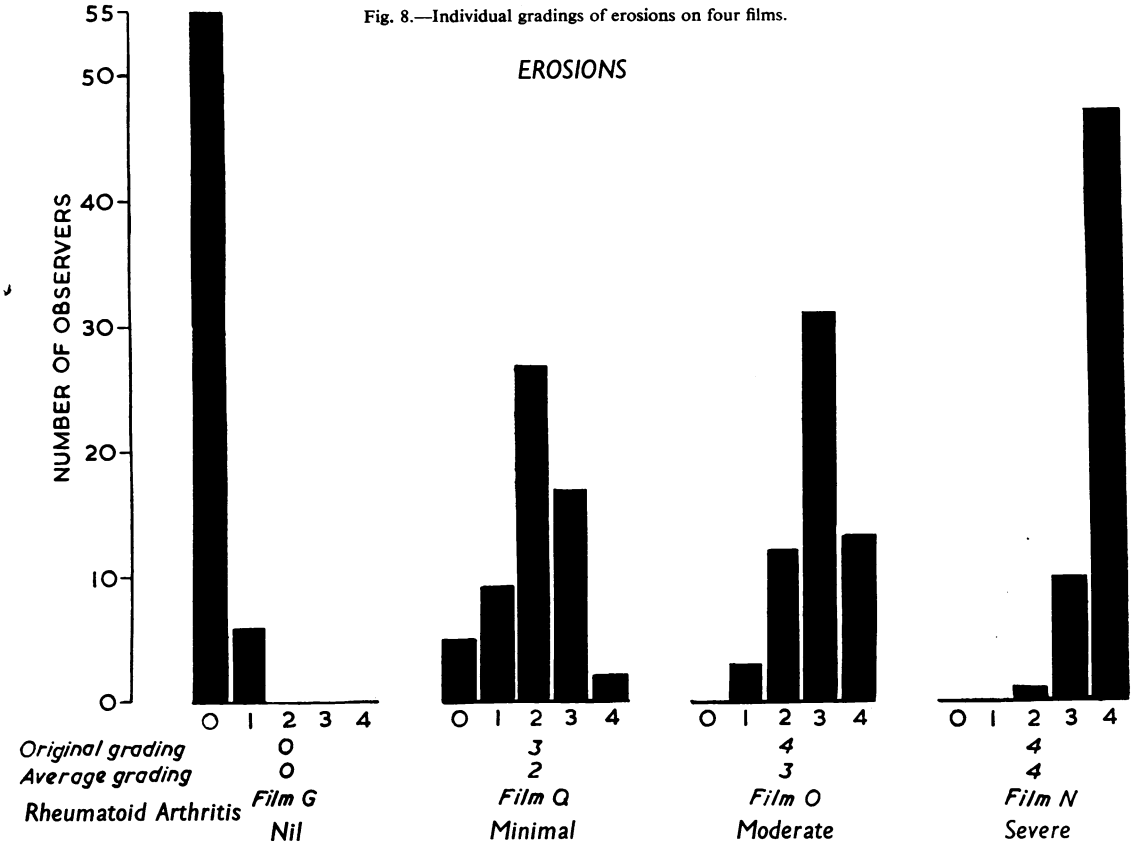


Fig. 8.—Individual gradings of erosions on four films.



change, a there was wider scatter of grading. On these films most observers tended to give the same grading for erosions as for rheumatoid arthritis.

The most striking differences of opinion were encountered on Film B from a cotton spinner who, according to the original British observers, had severe osteo-arthrosis in the distal and proximal interphalangeal joints of both hands but no evidence of rheumatoid arthritis (Fig. 9, overleaf). On this film only 33 per cent. of Congress observers gave a grading of no erosions, 25 per cent. grading it as 3 and 12 per cent. as 4. This high score for erosions, however, may have been due to misinterpretation of the original instructions, in which bone cysts and narrowing of joint space due to degenerative joint disease were to be excluded. In fact, only four observers gave this film a definite rheumatoid grading so that it would appear unlikely that rheumatoid erosions were intended by the majority.

Discussion

It is clear from the results of this inquiry that there is considerable unanimity of opinion on the radiological diagnosis of severe rheumatoid arthritis. Moreover, the complete absence of rheumatoid arthritis can be determined with some certainty. In rheumatoid arthritis of moderate severity, opinions are fairly unanimous as to the presence of rheumatoid arthritis and there is fair agreement as to grading, but where changes of rheumatoid arthritis are minimal wide differences of opinion are possible, not only on grading but even on the presence or absence of rheumatoid arthritis.

The original readings used by us in assessing the prevalence of rheumatoid arthritis in Britain reflect the mean values for all gradings at Toronto but tend to be higher all round. Such a finding is not unexpected. When reading survey films, the majority of which show no evidence of rheumatoid arthritis, there is inevitably a tendency for the

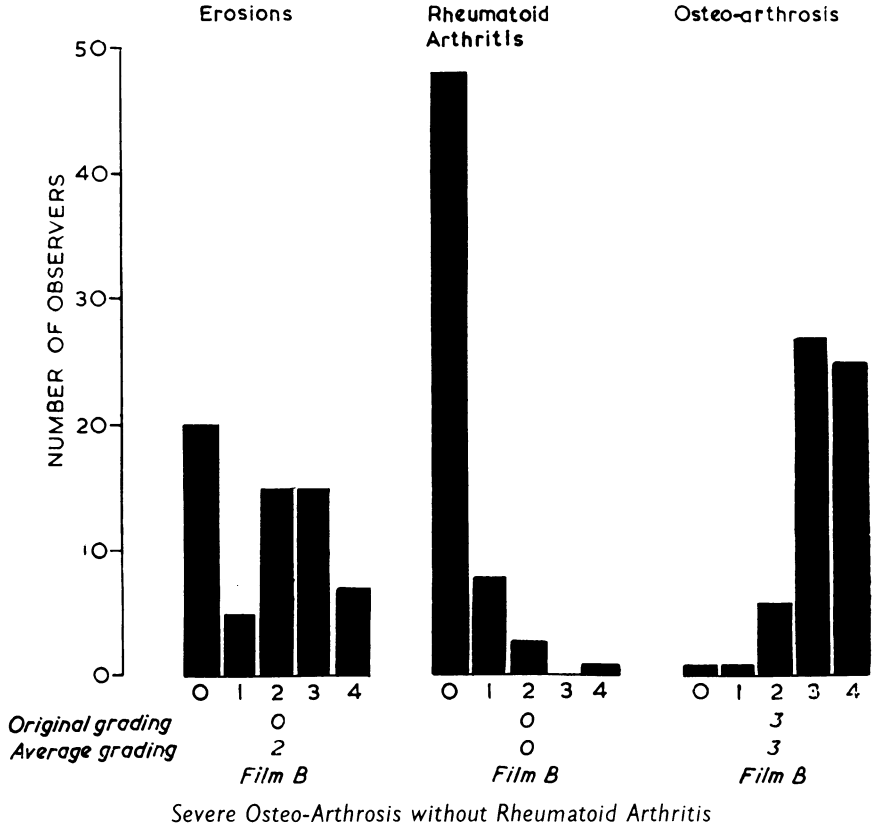


Fig. 9.—Individual gradings of erosions, rheumatoid arthritis, and osteo-arthrosis on Film B, showing severe osteo-arthrosis of the distal and proximal interphalangeal joints but no evidence of rheumatoid arthritis.

importance of relatively slight changes to be magnified and thus for grading to move upwards. This does not invalidate comparisons between respondents in a survey or even between different surveys, provided all the films are read by the same observer, but it may give a false and exaggerated impression of prevalence in surveys as a whole. For this reason it is suggested that standard films should be prepared for use by survey workers and others. As a preliminary, copies of Films G, Q, O, and N (Figs 1-4) from this inquiry could be made available on request. These give a useful measure of agreement on Grades 2, 3, and 4 in rheumatoid arthritis.

Summary

(1) The results of an inquiry held in Toronto during the Ninth International Congress of Rheumatic Diseases to determine suitable standards for the grading of radiological changes of rheumatoid arthritis are recorded.

(2) A satisfactory measure of agreement was found on the grading of rheumatoid arthritis and osteo-porosis, but there were wide differences of opinion on the interpretation of the term erosions. For all changes, agreement was greatest at either end of the scale with some scatter in the minimal and moderate grades.

(3) The original readings by workers engaged on population studies tended to have a higher grading than the average on all aspects except osteo-porosis on which they were slightly lower. No marked regional differences in grading were encountered.

(4) A series of films to illustrate average values for the three main grades of rheumatoid arthritis can now be made available to those taking part in comparative radiological studies.

Those taking part in the radiological quiz were:

United States of America: Drs. S. Y. Andelman, W. B. Andrus, P. J. Bilka, R. Brown, J. J. Bunim, G. M. Clark, S. Cobb, C. W. Denko, E. Dresner, M. W. Garry, J. D. Gowans, A. P. Hall, B. Hulbert, R. Irby, T. G. Kantor,

W. D. Kimler, H. F. Klinefelten, Jr., R. W. Lamont-Havers, C. McEwen, Anne B. McKusick, E. L. Persons, A. E. Phelps, V. K. Philips, C. Ragan, W. D. Robinson, H. Rosenfeld, E. F. Rosenberg, M. W. Ropes, A. Scherbel, K. Schmid, E. Scull, J. R. Shanahan, Smith, R. M. Stecher, L. T. Swaim, P. Thompson, L. L. Tormey, J. Vaughan, R. Wolpaw, M. Ziff;

England: Drs. J. F. Buchan, E. G. L. Bywaters, M. M. Dobson, O. Janus, A. G. S. Hill, W. E. Miall, W. A. Bourne;

France: Dr. J. Arlet;

Holland: Drs. J. J. de Blécourt, W. Hijmans;

Norway: Dr. V. Forbech;

Finland: Dr. V. A. Laine;

Spain: Dr. R. Moleres;

Canada: Drs. J. Durivage, J. P. Gofton, A. C. Kanaar, D. M. Mitchell;

Australia: Drs. J. B. Dick-Smith, S. Nelson;

Brazil: Dr. J. Houli.

REFERENCES

Medical Research Council and Nuffield Foundation Joint Committee on Clinical Trials of Cortisone, ACTH, and Other Therapeutic Measures in Chronic Rheumatic Diseases (1954). *Brit. med. J.*, **1**, 1223.

Kellgren, J. H., and Bier, F. (1956). *Annals of the Rheumatic Diseases*, **15**, 55.

—— and Lawrence, J. S. (1956). *Ibid.*, **15**, 1.

Évaluation radiologique de l'arthrite rhumatismale

RÉSUMÉ

(1) On présente les résultats d'une enquête menée durant le Neuvième Congrès International des Maladies Rhumatismales à Toronto pour déterminer les standards de classification, selon le degré, des altérations radiologiques dans l'arthrite rhumatismale.

(2) On est arrivé à un accord satisfaisant en ce qui concerne la classification de l'arthrite rhumatismale et de l'ostéoporose, mais il y avait de larges différences d'opinion dans l'interprétation du terme "érosion". Pour l'ensemble des altérations, l'accord était le plus complet aux deux bouts de l'échelle, avec quelques écarts dans les degrés minime et modéré.

(3) Les chercheurs engagés dans les enquêtes de dépistage, dans leurs interprétations originales des clichés offerts à la lecture, tendaient à une évaluation au dessus de la moyenne sous tous les aspects, sauf pour l'ostéoporose, où leur évaluation était légèrement en dessous. On n'observa aucune différence régionale appréciable dans la manière d'évaluer.

(4) Une série de clichés illustrant les valeurs moyens des trois degrés principaux d'arthrite rhumatismale est tenue maintenant à la disposition de ceux engagés dans des études radiologiques comparées.

Valoración radiológica de la artritis reumatoide

SUMARIO

(1) Se presentan los resultados de una encuesta durante el Noveno Congreso Internacional de Enfermedades Reumáticas en Toronto emprendida para determinar las normas de clasificación, según el grado, de las alteraciones radiológicas en la artritis reumatoide.

(2) Se llegó a un acuerdo satisfactorio respecto a la clasificación de la artritis reumatoide y de la ósteoporosis, pero hubo grandes diferencias de opinión en la interpretación del término "erosión". Para todas las alteraciones, el acuerdo fué más completo en ambos cabos de la escala, con algunas variaciones en los grados mínimo y moderado.

(3) Los investigadores de reumatismo en la población, cuyos clisés fueron empleados en la encuesta, tendían en sus interpretaciones originales a atribuir valores por encima de la media a todas las alteraciones, exceptuando la ósteoporosis, a la cual atribuían valores algo más bajos. No se observó diferencia regional apreciable en la manera de valorar.

(4) Una serie de clisés ilustrando los valores medios de los tres grados principales de artritis reumatoide está ahora a la disposición de investigadores empeñados en estudios radiológicos comparativos.